EXHIBIT 4

Y. CHOO, I. SANCHEZ-GARCIA & A. KLUG

NATURE 372 642-645, 1994

Exhibit 4

In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence

Yen Choo, Isidro Sánchez-García & Aaron Klug

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

A DNA-binding peptide comprising three zinc-fingers has been engineered to bind specifically to a unique nine-base-pair region of a *BCR–ABL* fusion oncogene in preference to the parent genomic sequences. Binding to the target oncogene in chromosomal DNA is possible in transformed cells in culture, and results in blockage of transcription. Consequently, murine cells rendered independent of growth factors by the action of the oncogene revert to factor dependence upon transient transfection with a vector expressing the peptide.

THE incorporation of proteins designed to recognize and bind specific DNA sequences into chimaeric transcription factors, recombinases and nucleases, for example, should have a wide range of applications. We have shown that zinc-finger minidomains can discriminate between closely related DNA triplets, and have proposed that they can be linked together to form domains for the specific recognition of longer DNA sequences[1,2]. One interesting possibility for the use of such protein domains is to target selectively genetic differences in pathogens or transformed cells. Here we report the first such application, in which we have built a protein which recognizes a specific DNA sequence both *in vitro* and *in vivo*.

There is a set of human leukaemias in which a reciprocal chromosomal translocation t(9;22)(q34;q11) results in a truncated chromosome 22, the Philadelphia chromosome (Ph^1)[3], encoding at the breakpoint a fusion of sequences from the c-*ABL* proto-oncogene[4] and the *BCR* gene[5]. In chronic myelogenous leukaemia (CML), the breakpoints usually occur in the first intron of the c-*ABL* gene and in the breakpoint cluster region of the *BCR* gene[6], and give rise to a gene product, $p210^{BCR\text{-}ABL}$, of relative molecular mass (M_r) 210,000 (ref. 7). Alternatively, in acute lymphoblastic leukaemia (ALL), the breakpoints usually occur in the first introns of both *BCR* and c-*ABL*[8], and result in a $p190^{BCR\text{-}ABL}$ gene product (Fig. 1)[9]. Facsimiles of these rearranged genes act as dominant transforming oncogenes in cell culture[10] and transgenic mice[11]. Like their genomic counterparts, the complementary DNAs bear a unique nucleotide sequence at the fusion point of the *BCR* and c-*ABL* genes, which can be recognized at the DNA level by a site-specific DNA-binding protein. We have designed such a protein to recognize the unique fusion site in the $p190^{BCR\text{-}ABL}$ cDNA. Our experiments are repeated on $p210^{BCR\text{-}ABL}$ cDNA which serves as a control. The exon fusions found in these cDNAs are obviously distinct from the breakpoints in the spontaneous genomic translocations, which occur in introns and are thought to be variable among patients. Hence although the design of peptides able to bind oncogenes has implications for cancer research, our primary aim here is to prove the principle of protein design, and to assess the feasibility of *in vivo* binding to chromosomal DNA in available model systems.

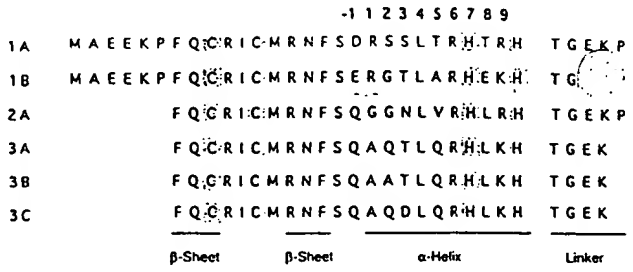
The DNA-binding proteins we create are composed of classical zinc-finger motifs[12–14]. These small motifs are ideal natural building blocks for *de novo* protein design because they function as independent modules[15], but can be connected by a well-known linker[16] to allow recognition of long, asymmetric DNA sequences. Lately it has been possible to isolate zinc-fingers that bind to given DNA triplets by selection from phage display libraries of randomized zinc-fingers[1,17,18]. The specificity of selected fingers is checked by a second selection technique called the 'binding-site signature', in which these fingers are used to screen libraries of randomized oligonucleotide-binding sites, thus identifying fingers that can specify a unique base triplet[2]. From these and other studies[19], elements of a recognition code have emerged which relate the amino-acid sequence of zinc-fingers to their cognate triplet.

The strategy we use in creating DNA-binding proteins combines phage display selection and rational design based on the available recognition rules. A 9-base-pair (bp) target sequence (GCA, GAA, GCC) for a three-zinc-finger peptide was chosen which spanned the fusion point of the $p190^{BCR\text{-}ABL}$ cDNA[8]. The three triplets forming this binding site were each used to screen a zinc-finger phage library over three rounds as described[1]. The selected fingers were then analysed by binding-site signatures to reveal their preferred triplet, and mutations to improve specificity were made to the finger selected for binding to GCA (Fig. 2). A phage display mini-library of putative *BCR–ABL*-binding three-finger proteins was cloned in fd phage, comprising six possible combinations of the six selected or designed fingers (1A, 1B; 2A; 3A, 3B and 3C) linked in the appropriate order. The mini-library was screened once with an oligonucleotide containing the 9 bp *BCR–ABL* target sequence, to select for tight binding clones over weak binders and background vector phage. Because the library was small, we did not include competitor DNA sequences for homologous regions of the genomic *BCR* and c-*ABL* genes, but instead checked the selected clones for their ability to discriminate. We found that although all the selected clones were

FIG. 1 Nucleotide sequences of the fusion point between *BCR* and *ABL* sequences in p190 cDNA, and of the corresponding exon boundaries in the *BCR* and c-*ABL* genes. Exon sequences are written in capital letters and introns in lower-case letters. Top, $p190^{BCR\text{-}ABL}$ cDNA; middle, *BCR* genomic sequence at junction of exon 1 and intron 1; bottom, *ABL*, genomic sequence at junction of intron 1 and exon 2 (ref. 8). The 9-bp target sequence in the $p190^{BCR\text{-}ABL}$ cDNA is underlined, as are the homologous sequences in genomic *BCR* and c-*ABL*.

BCR · ABL	TTC CAT GGA GAC GCA G AA GCC CTT CAG CGG CCA
BCR	TTC CAT GGA GAC GCA G gt gag ttc ctc acg cca
ABL	ccc ctt tct ctt cca g AA GCC CTT CAG CGG CCA

FIG. 2 Amino-acid sequences of zinc-fingers used in constructing the mini-library of putative *BCR–ABL* binders. Regions of secondary structure are underlined below the list; residue positions are given above, relative to the first position of the α-helix (position 1). Zinc-finger phages were selected from a library of 2.6×10^6 variants using three DNA-binding sites, each containing one of the triplets GCC, GAA or GCA[1]. Binding-site signatures (data not shown) indicate that fingers 1A and 1B specify the triplet GCC, finger 2A specifies GAA, and the fingers selected using the triplet GCA all prefer binding to GCT (ref. 2). These include finger 3A, the specificity of which we believed, on the basis of recognition rules, could be changed by a point mutation. Finger 3B, based on the selected finger 3A, but in which Gln at helical position +2 was altered to Ala, should be specific for GCA. Finger 3C is an alternative version of finger 3A, in which the recognition of C is mediated by Asp + 3 rather than by Thr + 3.



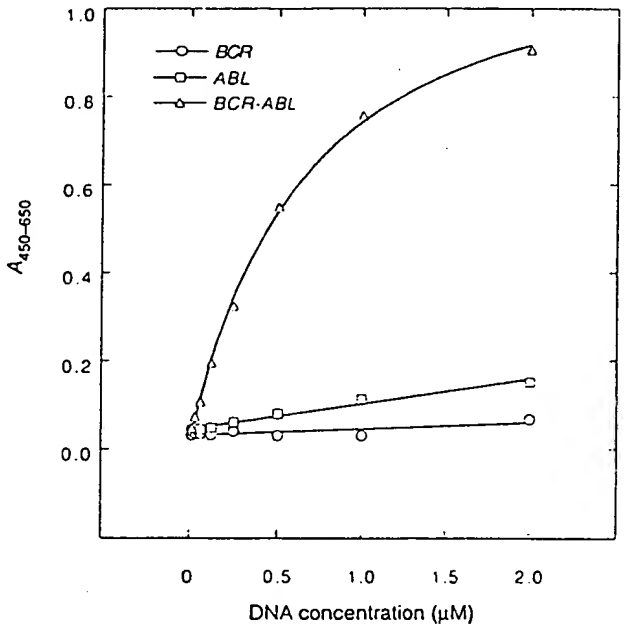


FIG. 3 Discrimination in the binding of the anti-BCR–ABL peptide to its $p190^{BCR-ABL}$ target site and to like regions of genomic *BCR* and c-*ABL*. The graph shows binding (measured as an absorbance between 450 and 650 nm) at various DNA concentrations. Binding reactions and complex detection by enzyme immunoassay were performed as described[2], and a full curve analysis was used in calculations of the K_d (ref. 16). The DNAs used were oligonucleotides spanning 9 bp either side of the fusion point in the cDNA or the exon boundaries. The anti-BCR–ABL peptide binds to its intended target site with $K_d = 6.2 \pm 0.4 \times 10^{-7}$ M, and is able to discriminate against genomic *BCR* and c-*ABL* sequences, although the latter differs by only one base pair in the bound 9-bp region.

able to bind the *BCR–ABL* target sequence and to discriminate between this and the genomic *BCR* sequence, only a subset could discriminate against the c-*ABL* sequence which, at the junction between intron 1 and exon 2, has an 8/9 bp homology to the *BCR–ABL* target sequence[8]. Sequencing of the discriminating clones revealed two types of selected peptide, one with the composition 1A-2A-3B and the other 1B-2A-3B. Thus both peptides carried the third finger (3B) which was specifically designed against the triplet GCA, but peptide 1A-2A-3B was able to bind to the *BCR–ABL* target sequence with higher affinity than was peptide 1B-2A-3B.

The peptide 1A-2A-3B, which we will refer to as the anti-BCR–ABL peptide, was used in further experiments. The anti-BCR–ABL peptide has an apparent equilibrium dissociation constant (K_d) of $6.2 \pm 0.4 \times 10^{-7}$ M for the $p190^{BCR-ABL}$ cDNA sequence *in vitro*, and discriminates against the similar sequences found in genomic *BCR* and c-*ABL* DNA by factors greater than an order of magnitude (Fig. 3). The measured dissociation constant is higher than that of three-finger peptides from naturally occurring proteins such as Sp1 (ref. 20) or Zif268 (ref. 21), which have K_d values in the range of 10^{-9} M, but rather is comparable to that of the two fingers from the *tramtrack* (*ttk*) protein[22]. However, the affinity of the anti-BCR–ABL peptide could be refined if necessary by site-directed mutations or by 'affinity maturation' of a phage display library[23].

Having established DNA discrimination *in vitro*, we wished to test whether the anti-BCR–ABL peptide was capable of site-specific DNA binding *in vivo*. The peptide was fused to the VP16 activation domain from herpes simplex virus[24] and used in transient transfection assays (Fig. 4) to drive production of a chloramphenicol acetyl transferase (CAT) reporter gene from a binding site upstream of the TATA box[25]. A thirtyfold increase in CAT activity was observed in cells cotransfected with reporter plasmid bearing copies of the $p190^{BCR-ABL}$ cDNA target site, compared to a barely detectable increase in cells cotransfected with reporter plasmid bearing copies of either the *BCR* or c-*ABL* semihomologous sequences. The selective stimulation of transcription indicates convincingly that highly site-specific DNA binding can occur *in vivo*. But although transient transfections assay binding to plasmid DNA, the true target site for this and most other DNA-binding proteins is in genomic DNA. This might well present significant problems, not least because this DNA is physically separated from the cytosol by the nuclear membrane, but also because it may be packaged within chromatin.

To study whether genomic targeting is possible, we made a construct in which the anti-BCR–ABL peptide was flanked at

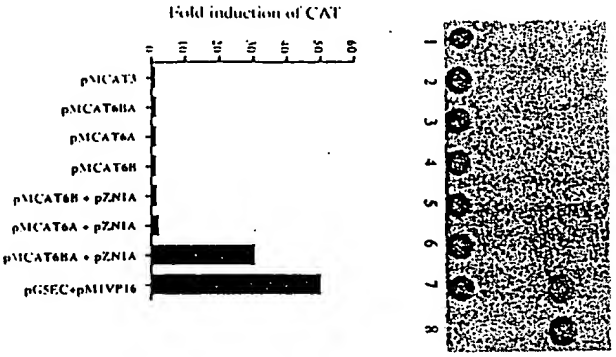


FIG. 4 Transactivation reporter assays using the anti-BCR–ABL peptide fused to the VP16 activation domain. C3H10T1/2 cells were transiently cotransfected with a CAT reporter plasmid and an anti-BCR–ABL/VP16 expression vector (pZN1A). The top panel shows the results of thin-layer chromatography of samples from different transfections, in which the fold induction of CAT activity relative to a sample in which reporter alonewas transfected (lane 1) is plotted on a histogram below. A specific (30-fold) increase in CAT activity was observed in cells cotransfected with reporter plasmid bearing copies of the $p190^{BCR-ABL}$ cDNA target site, indicating binding *in vivo*. The particular constructs used in different transfections are noted below the histogram.

METHODS. Reporter plasmids pMCAT6BA, pMCAT6A and pMCAT6B, were constructed by inserting 6 copies of the $p190^{BCR-ABL}$ target site (CGCAGAAGCC), the c-*ABL* second exon–intron junction sequence (TCCAGAAGCC), or the *BCR* first exon–intron junction sequence (CGCAGGTGAG), respectively, into pMCAT3 (ref. 33). The anti-BCR–ABL/VP16 expression vector was generated by inserting the in-frame fusion between the activation domain of herpes simplex virus VP16 (ref. 24) and the zinc-finger peptide in the pEF-BOS vector[34]. C3H10T1/2 cells were transiently cotransfected with 10 μg reporter plasmid and 10 μg expression vector. RSVL[35], which contains the Rous sarcoma virus long terminal repeat linked to luciferase, was used as an internal control to normalize for differences in transfection efficiency. Cells were transfected by the calcium phosphate precipitation method and CAT was assayed as described[36]. Plasmid pG5EC, which has five consensus 17-mer GAL4-binding sites upstream from the minimal promoter of the adenovirus E1b TATA box, and pM1VP16 vector, which encodes an in-frame fusion between the DNA-binding domain of GAL4 and the activation domain of herpes simplex virus VP16, were used as a positive control[37].

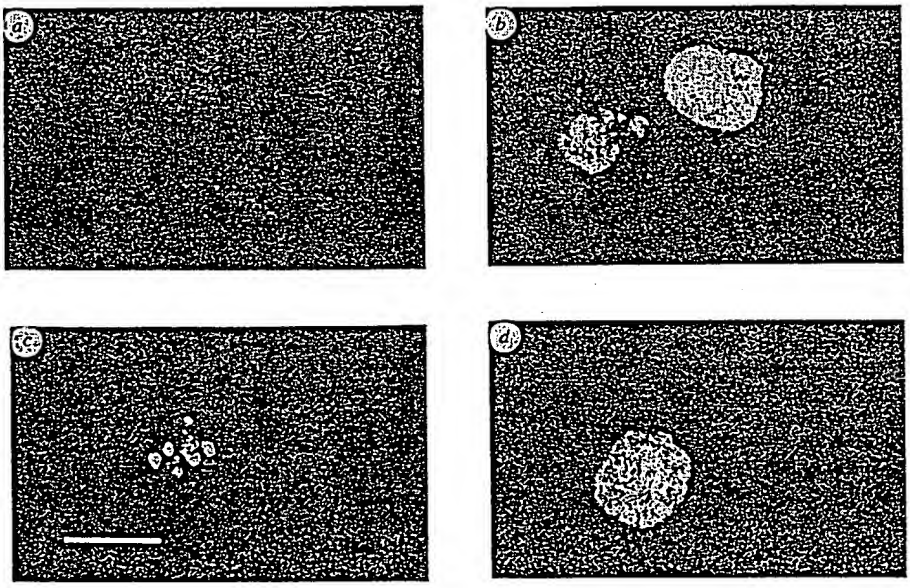


FIG. 5 Immunofluorescence microscopy in the absence of IL-3 of Ba/F3 + p190 and Ba/F3 + p210 cells transiently transfected with the anti-BCR–ABL expression vector and stained with the 9E10 antibody. The images show untransfected cells (*a*), and transfected cells in which the anti-BCR–ABL peptide is localized in the nucleus (*b*, *c* and *d*). 24 h after IL-3 withdrawal, transfected Ba/F3 + p190 cells show chromatin condensation and nuclear fragmentation into small apoptotic bodies (seen in various stages in *b* and in *c*, where the effect is more pronounced). In contrast, neither untransfected Ba/F3 + p190 cells (*a*) nor transfected Ba/F3 + p210 cells (*d*) show apoptosis. In *b* all three types of Ba/F3 + p190 cell are displayed: an untransfected cell (top right); a transfected but still intact cell (middle); and a transfected cell showing nuclear fragmentation (left).

METHODS. The anti-*BCR–ABL* expression vector was generated in the pEF-BOS vector[34], including an 11-amino-acid c-Myc epitope tag (EQKLISEEDLN) at the C-terminal end, recognizable by the 9E10 antibody[27] and the nuclear localization signal (PKKKRKV) of the large T antigen of SV40 virus[26] at the N-terminal end. Three glycine residues were introduced downstream of the nuclear localization signal as a spacer to ensure exposure of the nuclear leader from the folded molecule Ba/F3 cells were transfected with 25 μg of the anti-BCR-ABL expression construct tagged with the 9E10 c-Myc epitope as described[38]. Protein production was analysed 48 h later by immunofluorescence-labelling in the presence of IL-3, and also 24 h after subsequent withdrawal of IL-3. Cells were fixed in 4% (w/v) paraformaldehyde for 15 min, washed in PBS and permeabilized in methanol for 2 min. After blocking in 10% fetal calf serum in PBS for 30 min, the mouse 9E10 antibody was added. After 30-min incubation at room temperature, a fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG (SIGMA) was added and incubated for a further 30 min. Fluorescent cells were visualized using a confocal scanning microscope (magnification, ×90).

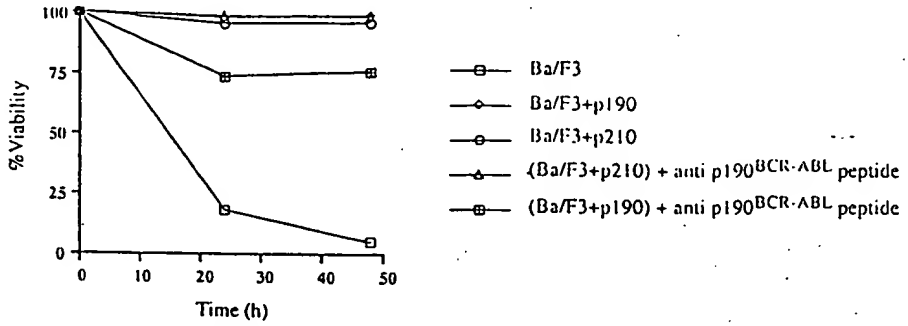


FIG. 6 Viability in the absence of IL-3 of transformed Ba/F3 cells transiently transfected with a vector expressing anti-BCR–ABL peptide. The Ba/F3 cell line is dependent on IL-3 for growth, but becomes IL-3-independent when stably transformed by $p190^{BCR-ABL}$ or $p210^{BCR-ABL}$ cDNA (ref. 10, and I.S.-G. *et al.*, manuscript in preparation). A proportion of Ba/F3 + p190 cells corresponding to the proportion transfected with the anti-BCR–ABL expression vector (Fig. 5) revert to IL-3 dependence, but similarly transfected Ba/F3 + p210 cells are unaffected.

the amino terminus with the nuclear localization signal from the large T antigen of SV40 virus[26], and at the carboxy terminus with an 11-amino-acid c-Myc epitope tag recognizable by the 9E10 antibody[27]. This construct was used transiently to transfect the interleukin-3 (IL-3)-dependent murine cell line Ba/F3 (ref. 28), or alternatively Ba/F3 + p190 and Ba/F3 + p210 cell lines previously made IL-3-independent by integrated plasmid constructs expressing either $p190^{BCR-ABL}$ or $p210^{BCR-ABL}$, respectively (I.S.G. *et al.*, manuscript in preparation). Immunofluorescence microscopy was used to visualize cells stained with 9E10 antibody followed by a secondary fluorescent conjugate (Fig. 5). The efficiency of transient transfection of all cell lines, measured as the proportion of immunofluorescent cells in the population, was 15–20%. The peptide is efficiently localized to the nucleus of all such transfected cells. In the presence of IL-3, all cell lines are viable 48 h after transfection. On the other hand, when IL-3 is subsequently withdrawn from cell culture, over 90% of the transfected Ba/F3 + p190 cells become apoptotic within 24 h (Fig. 5*b*, *c*). In contrast, all transfected Ba/F3 + p210 cells remain intact (Fig. 5*d*). The anti-BCR–ABL peptide is therefore able to restore an IL-3 dependence in the Ba/F3 + p190 cells which the oncogene had abolished.

In another set of experiments in which cell viability in tissue culture after IL-3 withdrawal is assayed by Tripan blue exclusion, 20% of Ba/F3 + p190 cells are found to have reverted to factor dependence and die within 24 h, whereas Ba/F3 + p210 cells are unaffected (Fig. 6). Northern blots of total cytoplasmic RNA from transfected Ba/F3 + p190 cells indicated levels of $p190^{BCR-ABL}$ messenger RNA reduced by 15–18% relative to untransfected cells. These results correlate well with the proportion of cells that have become transfected and which die in the absence of IL-3 (Figs 5 and 6). The reduction in mRNA level thus indicates a transcriptional block imposed by the sequence-specific binding of the peptide, which would obstruct the path of the polymerase.

Hence a DNA-binding protein designed to recognize a specific DNA sequence *in vitro*, is active *in vivo* where, directed to the nucleus by an appended localization signal, it can bind its target sequence in chromosomal DNA, causing a specific inhibition of transcription. The use of a blocking agent in this case to target intragenic sequences is reminiscent of antisense oligonucleotide- or ribozyme-based approaches to inhibiting the expression of selected genes[29]. Like antisense oligonucleotides, zinc-finger DNA-binding proteins can be tailored against genes altered by chromosomal translocations or point mutations. Also, like oligonucleotides that can be designed to repress transcription by triple-helix formation in homopurine–homopyrimidine promoters[30], DNA-binding proteins can bind to various unique regions outside genes. But by contrast to nucleic acids, proteins can direct gene expression by both up- or down-regulating the initiation of transcription when fused to activation[31] or repression domains[32].

By acting directly on any DNA, and by allowing fusion to a variety of protein effectors, tailored site-specific DNA-binding proteins have the potential to control expression of specific genes and to further the possibility of manipulating the genetic material itself, in medicine and research. □

Received 11 August; accepted 15 November 1994.

1. Choo, Y. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11163–11167 (1994).
2. Choo, Y. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11168–11172 (1994).
3. Nowell, P. C. & Hungerford, D. A. *Science* **132**, 1197–1200 (1960).
4. Bartram, C. R. *et al. Nature* **306**, 277–280 (1983).
5. Groffen, J. *et al. Cell* **36**, 93–99 (1984).
6. Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, E. *Nature* **315**, 550–554 (1985).
7. Konopka, J. B., Watanabe, S. M. & Witte, O. N. *Cell* **37**, 1035–1042 (1984).
8. Hermans, A. *et al. Cell* **51**, 33–40 (1987).
9. Kurzrock, R. *et al. Nature* **325**, 631–635 (1987).
10. Daley, G. Q. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9312–9316 (1988).
11. Heisterkamp, N. *et al. Nature* **344**, 251–253 (1990).
12. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
13. Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 99–102 (1988).
14. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. *Science* **245**, 635–637 (1989).
15. Pavletich, N. P. & Pabo, C. O. **252**, 809–817 (1991).
16. Choo, Y. & Klug, A. *Nucleic Acids Res.* **21**, 3341–3346 (1993).
17. Rebar, E. J. & Pabo, C. O. *Science* **263**, 671–673 (1994).
18. Jamieson, A. C., Kim, S.-H. & Wells, J. A. *Biochemistry* **33**, 5689–5695 (1994).
19. Desjarlais, J. R. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7345–7349 (1992).
20. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. *Cell* **51**, 1079–1090 (1987).
21. Christy, B. A., Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7857–7861 (1988).
22. Fairall, L., Harrison, S. D., Travers, A. A. & Rhodes, D. *J. molec. Biol.* **226**, 349–366 (1992).
23. Hawkins, R. E., Russell, S. J. & Winter, G. *J. molec. Biol.* **226**, 889–896 (1992).
24. Fields, S. *Methods* **5**, 116–124 (1993).
25. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
26. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. *Cell* 499–509 (1984).
27. Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. *Molec. cell. Biol.* **5**, 3610–3616 (1985).
28. Palacios, R. & Steinmetz, M. *Cell* **41**, 727–734 (1985).
29. Stein, C. A. & Cheng, Y. *Science* **261**, 1004–1012 (1993).
30. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J. & Hogan, M. E. *Science* **245**, 725–730 (1988).
31. Seipel, K., Georgiev, O. & Shaffner, W. *EMBO J.* **11**, 4961–4968 (1992).
32. Herschbach, B. M., Arnaud, M. B. & Johnson, A. D. *Nature* **370**, 309–311 (1994).
33. Luscher, B., Mitchell, P. J., Williams, T. & Tjian, R. *Genes Dev.* **3**, 1507–1517 (1989).
34. Mizushima, S. & Shigezaku, N. *Nucleic Acids Res.* **18**, 5322 (1990).
35. de Wet, J. R., Wood, K. V., de Luca, M., Helinski, D. R. & Subramani, S. *Molec. cell. Biol.* **7**, 725–737 (1987).
36. Sanchez-Garcia, I., Osada, H., Forster, A. & Rabbitts, T. H. *EMBO J.* **12**, 4243–4250 (1993).
37. Sadowski, I., Brendan, B., Broad, P. & Hollis, M. *Gene* **118**, 137–141 (1992).
38. Sanchez-Garcia, I. & Rabbitts, T. H. *Proc. natn. Acad. Sci. U.S.A.* **91**, 7869–7873 (1994).

ACKNOWLEDGEMENTS. We thank T. Rabbitts for transformed Ba/F3 cell lines and for advice, J. Girdlestone for the pMCAT3 vector, I. Sadowski for plasmids pG5EC and pM1VP16 and S. Munro for the 9E10 antibody. Y.C. is supported by a British Council Fellowship and a grant from the MRC. I.S.-G. is supported by European Community fellowship.